

Nam D. Nguyen

PHD CANDIDATE IN COMPUTER SCIENCE · MACHINE LEARNING RESEARCHER · BIOINFORMATICIAN

2316 Badger Parkway, Madison, WI 53713, USA

☎ +1 (631) 561-7343 | ✉ ndnguyen@cs.stonybrook.edu | 🏠 namtk.github.io | 📷 namtk | 📺 namtrk

Summary

I focus on the tasks of machine learning on graphs and manifolds with applications currently on understanding molecular mechanisms and improving genotype-phenotype predictions in complex biological systems. With my expertise in Python/PyTorch and years of experience of software engineering, I also desire to develop large-scale data analytic and scalable machine learning systems for real-world applications.

Research Interests

Machine Learning Multiview Learning, Representation Learning, Geometric Deep Learning, Explainable AI, Statistical Learning Theory
Network Science Link Prediction, Network Diffusion, Network Embedding, Network Curvatures, Social & Biological Network Analysis
Bioinformatics Single-cell Multiomics, Gene Expression & Regulation, Precision Medicine, Computational Genomics

Education

Stony Brook University

Stony Brook, NY, USA

PH.D. IN COMPUTER SCIENCE

Aug. 2016 - Exp. June 2021

- Thesis Proposal: “Interpretable Multiview Learning for Understanding Functional Multiomics” · GPA: 3.84/4.00
- Thesis committee: Drs. Daifeng Wang (advisor), Fusheng Wang (committee chair), Steven Skiena
- Relevant coursework: Machine Learning & Convex Optimization, Data Science, Analysis of Algorithms, Translational Bioinformatics

Hanoi University of Science and Technology

Hanoi, Vietnam

B.ENG. IN COMPUTER SCIENCE

Aug. 2005 - Aug. 2010

- Thesis Title: “Integrating Security into Software Engineering, Using Security Patterns and SecureUML” · GPA: 3.16/4.00
- Relevant coursework: Numerical Methods, Digital Signal Processing, Artificial Intelligence, Software Architecture & Reuse, Information Security

Experience

Daifeng Wang Laboratory, Waisman Center, University of Wisconsin-Madison

Madison, WI, USA

RESEARCH FELLOW

Sep. 2019 – Current

- Design novel machine learning algorithms for problems in functional genomics, cancers, & brain disorders (NIH-funded projects)
- Implement & deploy these algorithms, resulting in their high empirical performance, & maintain them as Python/R packages on GitHub
- Formulate a theoretical framework for multiview learning, MV-ERM, to unify & analyze state-of-the-art multiview learning algorithms
- Solve the trade-off between parametric & nonlinear manifold alignment by designing a variant of stochastic gradient descent on a Riemannian space to train a pair of coupled deep neural nets
- Survey state-of-the-art quantum machine learning algorithms for potential applications in genomics
- Collect & analyze single-cell (scRNA-Seq) data to identify the genetic risk of variation associated with neuropsychiatric symptoms (in collaboration with Icahn School of Medicine at Mount Sinai)

Daifeng Wang Laboratory, Stony Brook University (SUNY)

Stony Brook, NY, USA

RESEARCH ASSISTANT

May 2017 - Aug 2019

- Designed, implemented, & deployed algorithms using manifold alignment to reveal functional links between gene networks
- Explored the capabilities of manifold alignment in the settings of multiview learning
- Designed & implemented computational pipelines to analyze RNA-Seq data for bladder cancer research (NIH-funded project)
- Led a team in 2 prediction competitions on Kaggle (top 22%) & a project of training the machine to distinguish good from poor singing

Department of Biology, Brookhaven National Laboratory

Brookhaven, NY, USA

RESEARCH ASSISTANT

May 2017 - May 2018

- Developed computational pipeline to collect & analyze RNA-Seq data of algae for the project “Large-Scale Comparative Regulatory Network Analysis in Photosynthetic Organisms” (DOE-funded project)

Software Engineering Lab., Pohang University of Science and Technology (POSTECH)

Pohang, S. Korea

RESEARCHER & SOFTWARE ENGINEER

Mar. 2011 - May 2013

- Researched feature-oriented product line engineering & model checking
- Specified & implemented a set of rules to automatically verify consistency between a feature model & other requirements models
- Investigated the complex network structures (e.g., small world, scale-free) & formulated a set of software metrics for large-scale software systems

Graphics & Virtual Reality Lab., Vietnam Academy of Science and Technology

Hanoi, Vietnam

RESEARCH INTERN

Jun. 2007 - Aug. 2007

- Surveyed & applied the wavelet transform to image denoising & compression; implemented a prototype in MATLAB

Publications

JOURNALS

Nguyen, N. D., Jin, T., & Wang, D. (2020). *Varmole: A biologically drop-connect deep neural network model for prioritizing disease risk variants and genes*. *Bioinformatics*, , btaa866

Jin*, T., **Nguyen***, N. D., Talos, F., & Wang, D. (2020). *ECMarker: Interpretable machine learning model identifies gene expression biomarkers predicting clinical outcomes and reveals molecular mechanisms of human disease in early stages*. *Bioinformatics*, , btaa935 (*joint first authorship)

Nguyen, N. D., & Wang, D. (2020). *Multiview learning for understanding functional multiomics*. *PLOS Computational Biology*, **16**(4), e1007677.

Nguyen, N. D., Blaby, I. K., & Wang, D. (2019). *ManiNetCluster: A novel manifold learning approach to reveal the functional links between gene networks*. *BMC Genomics*, **20**(12), 1–14.

SELECTED CONFERENCE TALKS & PRESENTATIONS

Nguyen, N. D., Jin, T., & Wang, D. (2020). *Varmole: A biologically drop-connect deep neural network model for prioritizing disease risk variants and genes*. *The 13th annual RECOMB/ISCB conference on regulatory & systems genomics with DREAM challenges (RECOMB RSGDREAM)*.

Nguyen, N. D., & Wang, D. (2020). *Multiview learning for understanding functional multiomics*. *The 28th conference on intelligent systems for molecular biology (ISMB)*.

Nguyen, N. D., Blaby, I., & Wang, D. (2019). *ManiNetCluster: A novel manifold learning approach to reveal the functional links between gene networks*. *The international conference on intelligent biology and medicine (ICIBM)*. Columbus, OH, USA. (one of Top 7 papers out of 105 submissions)

Nguyen, N. D., Blaby, I., & Wang, D. (2018). *A manifold learning based approach to reveal the functional linkages across multiple gene networks*. *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics (ACM-BCB)*, 514–514. Washington, D.C., USA. (award paper)

UNDER REVIEW/IN PREPARATION

Nguyen, N. D., Huang, J., & Wang, D. (2021). *deepManReg: A deep manifold-regularized learning model for improving phenotype prediction from multi-modal data*. *bioRxiv 2021.01.28.428715*. **Under Review**.

Nguyen, N. D., & Wang, D. (2021). *Scalable Deep Manifold Alignment via Stiefel Manifold Optimization & Nystrom Method*. **In Preparation**.

Honors & Awards

2021	Ray and Stephanie Lane Fellowship , School of Computer Science, Carnegie Mellon University	USA
2021	DAAD PostDoc-Net-AI Fellowship (~10% acceptance rate), Federal Ministry of Education and Research	Germany
2019	NSF Travel Award , The US National Science Foundation (ICIBM 2019)	Columbus, OH, USA
2018	SIGBio Best Poster Award , ACM SIGBio (ACM-BCB 2018)	USA
2016-17	VEF Doctoral Fellowship , The US National Academy of Sciences	USA
2011-12	POSTECH Research Scholarship , National Research Foundation of Korea	S. Korea
2010	Scholarship for Excellent Students , Ministry of Education and Training of Vietnam	Vietnam

Teaching & Services

2018	Mentor , Computer Science and Informatics Summer Research Experience Program	Stony Brook Univ.
2018	Guest Lecturer , BMI511 - Translational Bioinformatics	Stony Brook Univ.
2016-17	Teaching Assitant , CSE101 - Computer Science Principles & CSE114 - Intro. to Object Oriented Programming	Stony Brook Univ.
2018-19	Member , ISCB (International Society for Computational Biology), ACM (Association for Computing Machinery)	
2011	Organizing Volunteer , 12th International Conference on Software Reuse (ICSR 2011)	Pohang, S. Korea

Skills & Certifications

Programming	Python/PyTorch, R (5-year experience); MATLAB (3-year experience); C++, Java (2-year experience)
Others	UNIX Shell, AWK, Vim, Git, LaTeX, Markdown & RMarkdown, Jupyter Notebook, Docker/Singularity
Bioinformatics	Next-generation Sequencing Analysis (RNA-Seq), Differential Gene Expression & Enrichment Analysis, WGCNA
Mathematics	Linear Algebra, Probability & Statistics, Multivariate Calculus, Convex Optimization
Coursera Certificates	Social Network Analysis; Game Theory; Model Thinking
IBM Certified	Solution Designer - Object-Oriented Analysis & Design; Application Developer; Database Associate

References

Daifeng Wang, PhD Assistant Professor; University of Wisconsin-Madison; Email: daifeng.wang@wisc.edu

Ian K. Blaby, PhD Group Lead/Scientist; US Department of Energy, Lawrence Berkeley National Laboratory; Email: ikblaby@lbl.gov